

#2



OIPE

RAW SEQUENCE LISTING

DATE: 02/14/2002

PATENT APPLICATION: US/10/054,935

TIME: 16:05:09

Input Set : A:\16U 107 R1.ST25.txt

Output Set: N:\CRF3\02142002\J054935.raw

3 <110> APPLICANT: OriGene Technologies, Inc
 5 <120> TITLE OF INVENTION: BREAST CANCER TRANSCRIPTION FACTOR GENE AND USES
 7 <130> FILE REFERENCE: 16U 107 R1
 C--> 9 <140> CURRENT APPLICATION NUMBER: US/10/054,935
 C--> 9 <141> CURRENT FILING DATE: 2002-01-25
 9 <160> NUMBER OF SEQ ID NOS: 4
 11 <170> SOFTWARE: PatentIn version 3.1
 13 <210> SEQ ID NO: 1
 14 <211> LENGTH: 4372
 15 <212> TYPE: DNA
 16 <213> ORGANISM: Homo sapiens
 18 <220> FEATURE:
 19 <221> NAME/KEY: CDS
 20 <222> LOCATION: (78)..(1922)
 21 <223> OTHER INFORMATION:
 24 <400> SEQUENCE: 1
 25 cagtcctcga cccccgcac ctgcgccctt cccaccccc tcctccgcct cgggtgcccg 60
 27 cgctgctccg gaccact atg acc atg aga tcc gcg gtg ttc aag gcg gcc 110
 28 Met Thr Met Arg Ser Ala Val Phe Lys Ala Ala
 29 1 5 10
 31 gcg gcc cct gcc ggc aat cct gag cag cga ctg gac tac gag cgg 158
 32 Ala Ala Pro Ala Gly Gly Asn Pro Glu Gln Arg Leu Asp Tyr Glu Arg
 33 15 20 25
 35 gct gcg gcg ctg ggc ggg ccc gag gac gag cct ggg gcg gcc gaa gcc 206
 36 Ala Ala Ala Leu Gly Gly Pro Glu Asp Glu Pro Gly Ala Ala Glu Ala
 37 30 35 40
 39 cac ttc ctc ccc cgg cac cgt aag ctc aag gag ccg ggg ccc ccg ctg 254
 40 His Phe Leu Pro Arg His Arg Lys Leu Lys Glu Pro Gly Pro Pro Leu
 41 45 50 55
 43 gcc tcc tcc cag ggc ggg agc ccc gcg cct tcc ccg gcc ggc tgc ggc 302
 44 Ala Ser Ser Gln Gly Gly Ser Pro Ala Pro Ser Pro Ala Gly Cys Gly
 45 60 65 70 75
 47 ggc aag ggc cgg ggc ttg tta ctc ccg gcc ggg gcg gcc ccc ggg cag 350
 48 Gly Lys Gly Arg Gly Leu Leu Leu Pro Ala Gly Ala Ala Pro Gly Gln
 49 80 85 90
 51 cag gaa gag agc tgg ggc ggt tcg gtg ccc ttg ccc tgt ccg ccc ccg 398
 52 Gln Glu Glu Ser Trp Gly Gly Ser Val Pro Leu Pro Cys Pro Pro Pro
 53 95 100 105
 55 gcc acc aag caa gcc ggc att ggg ggg gag cct gcc gca gcc gga gcc 446
 56 Ala Thr Lys Gln Ala Gly Ile Gly Gly Glu Pro Ala Ala Ala Gly Ala
 57 110 115 120
 59 ggc tgc agc ccc cgg ccc aag tat cag gcg gtg ctg ccc att cag acg 494
 *60 Gly Cys Ser Pro Arg Pro Lys Tyr Gln Ala Val Leu Pro Ile Gln Thr

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/054,935

DATE: 02/14/2002

TIME: 16:05:09

Input Set : A:\16U 107 R1.ST25.txt

Output Set: N:\CRF3\02142002\J054935.raw

61	125	130	135	
63	ggc tct ctc gtg gcg gcg gcc aaa gag cct acg ccc tgg gct ggg gac	542		
64	Gly Ser Leu Val Ala Ala Ala Lys Glu Pro Thr Pro Trp Ala Gly Asp			
65	140	145	150	155
67	aag ggt ggg gcg gcc tcc ccc gct gcc acc gcc tcg gac ccg gcg gga	590		
68	Lys Gly Gly Ala Ala Ser Pro Ala Ala Thr Ala Ser Asp Pro Ala Gly			
69		160	165	170
71	ccc cca cca cta cct ctg ccc ggg ccg cca ccc ctc gcg ccc acc gcc	638		
72	Pro Pro Pro Leu Pro Leu Pro Gly Pro Pro Pro Leu Ala Pro Thr Ala			
73		175	180	185
75	acc gcc ggg acc ctg gcg gcc agc gag ggc aga tgg aag agt atg agg	686		
76	Thr Ala Gly Thr Leu Ala Ala Ser Glu Gly Arg Trp Lys Ser Met Arg			
77		190	195	200
79	aag agc cct ctc ggg ggt ggt ggc ggc tcg gga gcc tcc agt cag gcc	734		
80	Lys Ser Pro Leu Gly Gly Gly Gly Gly Ser Gly Ala Ser Ser Gln Ala			
81		205	210	215
83	gcc tgc ctc aaa cag atc ctt ctg ctg caa ttg gac ctc atc gaa cag	782		
84	Ala Cys Leu Lys Gln Ile Leu Leu Leu Gln Leu Asp Leu Ile Glu Gln			
85	220	225	230	235
87	cag cag cag cag ctg cag gcc aag gaa aag gag atc gag gag ctg aag	830		
88	Gln Gln Gln Gln Leu Gln Ala Lys Glu Lys Glu Ile Glu Glu Leu Lys			
89		240	245	250
91	tca gag aga gac acg ctc ctt gct cgg att gaa cgt atg gaa agg cgg	878		
92	Ser Glu Arg Asp Thr Leu Leu Ala Arg Ile Glu Arg Met Glu Arg Arg			
93		255	260	265
95	atg cag ctg gta aag aag gat aac gag aaa gaa agg cac aag ctg ttt	926		
96	Met Gln Leu Val Lys Lys Asp Asn Glu Lys Glu Arg His Lys Leu Phe			
97		270	275	280
99	cag ggc tat gaa act gaa gag aga gag gaa aca gag cta tct gag aaa	974		
100	Gln Gly Tyr Glu Thr Glu Glu Arg Glu Glu Thr Glu Leu Ser Glu Lys			
101		285	290	295
103	att aaa ctg gag tgc cag ccg gag ctt tcc gag aca tcc cag act ctg	1022		
104	Ile Lys Leu Glu Cys Gln Pro Glu Leu Ser Glu Thr Ser Gln Thr Leu			
105	300	305	310	315
107	cct ccc aag ccc ttc tca tgt ggg cgg agt gga aag gga cat aaa agg	1070		
108	Pro Pro Lys Pro Phe Ser Cys Gly Arg Ser Gly Lys Gly His Lys Arg			
109		320	325	330
111	aaa tcc cca ttt gga agt aca gaa aga aag act cct gtt aaa aag ctg	1118		
112	Lys Ser Pro Phe Gly Ser Thr Glu Arg Lys Thr Pro Val Lys Lys Leu			
113		335	340	345
115	gct cct gaa ttt tca aaa gtc aaa aca aaa act cct aag cac tct cct	1166		
116	Ala Pro Glu Phe Ser Lys Val Lys Thr Lys Thr Pro Lys His Ser Pro			
117		350	355	360
119	att aaa gag gaa ccc tgt ggt tcc tta tct gaa act gtt tgt aaa cgt	1214		
120	Ile Lys Glu Glu Pro Cys Gly Ser Leu Ser Glu Thr Val Cys Lys Arg			
121		365	370	375
123	gaa ttg agg agc caa gaa acc cca gaa aag ccc cgg tct tca gtg gac	1262		
124	Glu Leu Arg Ser Gln Glu Thr Pro Glu Lys Pro Arg Ser Ser Val Asp			
125	380	385	390	395

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/054,935

DATE: 02/14/2002

TIME: 16:05:09

Input Set : A:\16U 107 R1.ST25.txt

Output Set: N:\CRF3\02142002\J054935.raw

```

127 acc cca cca aga ctc tcc act ccc caa aag gga ccc agc acc cat ccc 1310
128 Thr Pro Pro Arg Leu Ser Thr Pro Gln Lys Gly Pro Ser Thr His Pro
129 400 405 410
131 aag gag aaa gcc ttc tca agt gag ata gaa gat ttg ccg tac ctt tcc 1358
132 Lys Glu Lys Ala Phe Ser Ser Glu Ile Glu Asp Leu Pro Tyr Leu Ser
133 415 420 425
135 acc aca gaa atg tat ttg tgt cgt tgg cac cag cct ccc cca tca ccg 1406
136 Thr Thr Glu Met Tyr Leu Cys Arg Trp His Gln Pro Pro Ser Pro
137 430 435 440
139 tta cca tta cgg gaa tcc tct cca aag aag gag gag act gta gca agg 1454
140 Leu Pro Leu Arg Glu Ser Ser Pro Lys Lys Glu Glu Thr Val Ala Arg
141 445 450 455
143 tgt ctg atg cca tca agt gtt gca gga gaa act tca gtc ttg gct gtt 1502
144 Cys Leu Met Pro Ser Ser Val Ala Gly Glu Thr Ser Val Leu Ala Val
145 460 465 470 475
147 cct tct tgg agg gac cac tca gta gag cct cta agg gac cca aat cct 1550
148 Pro Ser Trp Arg Asp His Ser Val Glu Pro Leu Arg Asp Pro Asn Pro
149 480 485 490
151 tca gac ctt ttg gag aac ctg gat gac agt gtg ttt tcg aag cgg cat 1598
152 Ser Asp Leu Leu Glu Asn Leu Asp Asp Ser Val Phe Ser Lys Arg His
153 495 500 505
155 gca aaa ctg gag ctg gat gag aag aga agg aaa aga tgg gat att cag 1646
156 Ala Lys Leu Glu Leu Asp Glu Lys Arg Arg Lys Arg Trp Asp Ile Gln
157 510 515 520
159 agg atc agg gaa caa aga att tta cag cga ctg cag ctc aga atg tat 1694
160 Arg Ile Arg Glu Gln Arg Ile Leu Gln Arg Leu Gln Leu Arg Met Tyr
161 525 530 535
163 aaa aag aaa gga att cag gaa tct gag cct gag gtt acc tca ttt ttc 1742
164 Lys Lys Lys Gly Ile Gln Glu Ser Glu Pro Glu Val Thr Ser Phe Phe
165 540 545 550 555
167 cct gag cca gat gat gtt gaa agt ttg atg att acc ccc ttc ttg cct 1790
168 Pro Glu Pro Asp Asp Val Glu Ser Leu Met Ile Thr Pro Phe Leu Pro
169 560 565 570
171 gtt gta gca ttt gga cga cca tta cca aaa tta act cca cag aat ttt 1838
172 Val Val Ala Phe Gly Arg Pro Leu Pro Lys Leu Thr Pro Gln Asn Phe
173 575 580 585
175 gag cta ccc tgg ttg gat gag cgt agc cga tgc aga ttg gag atc cag 1886
176 Glu Leu Pro Trp Leu Asp Glu Arg Ser Arg Cys Arg Leu Glu Ile Gln
177 590 595 600
179 aag aag caa aca cct cac cgg acg tgt agg aaa tag ctgtgctggc 1932
180 Lys Lys Gln Thr Pro His Arg Thr Cys Arg Lys
181 605 610
183 aagaaccctg tcttcagata gttgtagcat gccattcccc agagtggcag agacctgtat 1992
185 atgtgacctt tgtcctcaca tatgttatca ctgcgtgata ataccctttc atacttcctt 2052
187 gactttgttt tcattactct gatttcacaa aaactctttc attcggctaa ttgtgagtta 2112
189 tggagggtga ttgggatttc ttttcccttt ttgggaaat gggctctcaa gctaaagcta 2172
191 taggatggca gattcagaag tttcaggggt ctgtttctat acatttgcct atgttaaagg 2232
193 ggtaaaaagg ctctcttcat tagacatgtg gaagatgaag cagccccttc ctttagagct 2292
195 gtgcctgcat ggcactcttc tcaccctggt acaccctcct tatagtgggt atagtgattt 2352

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/054,935

DATE: 02/14/2002

TIME: 16:05:09

Input Set : A:\16U 107 R1.ST25.txt

Output Set: N:\CRF3\02142002\J054935.raw

```

197 ttaaccctaa aataaaacaa acaacctcac catgagcttt aggaccagaa gaggaatgac 2412
199 aagtgaagcg atgaagcaag ccatcttcac agagtagaaa agacatcgga gagttggtag 2472
201 ataactgtct gaaaagatag ttgttcattt gaaactatto tgtgatacag tcatgtggga 2532
203 agggatgttt ggctgtgatt attttttccag ttaatggata acaatttctt tactgtctaa 2592
205 aaacaaaaat ctttgaaaaa gaaagtgggg atggttagtt tcagaacaag ttacagctgt 2652
207 aaacaaaagc acttagtatt tgggatggca tgccaaaacc tgtataaatg tccttgatc 2712
209 acatcacttc tcaagtattc cttcattggg cttcatcctt ttagcagaac tcttggtggt 2772
211 gggatagaga cttagggagg gtagggggag agtgtggaaa taggtgcttc ctttggtggt 2832
213 caaatgtcta catcttgaaa caaacagatg tacctaatag gcttctccat tcactttgta 2892
215 aaaataattt gtatgtgtac catcttggtc cttcccctc cgtttttgtt aaaatatcag 2952
217 gatagcactc ccaggccact ttggtctcag tgtaagatcc ctattaacta tctgaaagga 3012
219 aaatagagcc aagacctctg gtctcaaata tataggaatt gcctttcttt agtcttcagg 3072
221 actatttgtt gaaaacaagt aggggtctaa tctcctagaa ggtaggggct tttatcctta 3132
223 aagagaatat gtcccagat tattagcact tttagaggag aagccaaggt atgtaggggtg 3192
225 tgtggctggc ccatcagtg agcacgaaga gagaatggga taccattgtg ggaagagaag 3252
227 aaaagttcct caggggcctc ccaactgctaa agttttttgt gagatgttga tctgtgcttc 3312
229 ctggatttga cttttaaagg aattattctg gcagcacatg tagtattctt ggatgatctt 3372
231 gctgctctta tttctccttt tgtgtgtgtg tgtgtgtgtg tgtggctatg ggttttcatt 3432
233 tgtaactcca tctgcttagg agagtgggct cctataaagg gaacctgctg taaacttcat 3492
235 tgcagcaagg atgtagagag aaataggact taattccact aggggctctc atctcacacc 3552
237 tggagga gatttctaga aaaactgggc cagattttct ttgttctcca tcattttaat 3612
239 tgggct gttcagtttt cttactctta cctatgtgat atttcttctg aacgtgtcca 3672
241 taaaaa agaccctaac agtgtctctt gaatttggtt tttgatccct cagtttcttc 3732
243 ttcag catgtgtcgg gttcctaatt ttgggtatga gttagcaaat ttaaccattg 3792
245 tgtttgtgcc ctaccagggg gactcccag tttctgactt gaagtagact gagaagaatc 3852
247 cagagggtgc tatctggcca gatttaagta gattctatct ccttggttct cctctcctc 3912
249 gaggacctct tattttattg tcccctcttc taggttaatt ctcctttgat ttgactttgt 3972
251 tgaagaaggag gttggacagt agattagcaa agttccaagt gcaaaattac agtgtgttag 4032
253 agtgtggggg gaaaattagt cttatttttc cctacatggg atacaacact gtgaattcaa 4092
255 tcttcaactg aaggccctgc agttctccta aaacatagtt gtttgttttt ctttaacaaa 4152
257 gtttaagcta gtgttaataa attaaaaaaa attgcttgct tgtctacttc agctttgttt 4212
259 tatgccatt tcatattgtt gtctgtgttg taattcataa cttttgatac catttctgat 4272
261 gtgtaaaatt gggtgtcttg taaatatctt ataaagagtt caattgtaaa taaactattg 4332
263 tggctgttaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4372

```

266 <210> SEQ ID NO: 2

267 <211> LENGTH: 614

268 <212> TYPE: PRT

269 <213> ORGANISM: Homo sapiens

271 <400> SEQUENCE: 2

```

273 Met Thr Met Arg Ser Ala Val Phe Lys Ala Ala Ala Ala Pro Ala Gly
274 1 5 10 15
277 Gly Asn Pro Glu Gln Arg Leu Asp Tyr Glu Arg Ala Ala Ala Leu Gly
278 20 25 30
281 Gly Pro Glu Asp Glu Pro Gly Ala Ala Glu Ala His Phe Leu Pro Arg
282 35 40 45
285 His Arg Lys Leu Lys Glu Pro Gly Pro Pro Leu Ala Ser Ser Gln Gly
286 50 55 60
289 Gly Ser Pro Ala Pro Ser Pro Ala Gly Cys Gly Gly Lys Gly Arg Gly
290 65 70 75 80

```

RAW SEQUENCE LISTING

DATE: 02/14/2002

PATENT APPLICATION: US/10/054,935

TIME: 16:05:09

Input Set : A:\16U 107 R1.ST25.txt

Output Set: N:\CRF3\02142002\J054935.raw

```

293 Leu Leu Leu Pro Ala Gly Ala Ala Pro Gly Gln Gln Glu Glu Ser Trp
294      85      90      95
297 Gly Gly Ser Val Pro Leu Pro Cys Pro Pro Pro Ala Thr Lys Gln Ala
298      100      105      110
301 Gly Ile Gly Gly Glu Pro Ala Ala Ala Gly Ala Gly Cys Ser Pro Arg
302      115      120      125
305 Pro Lys Tyr Gln Ala Val Leu Pro Ile Gln Thr Gly Ser Leu Val Ala
306      130      135      140
309 Ala Ala Lys Glu Pro Thr Pro Trp Ala Gly Asp Lys Gly Gly Ala Ala
310 145      150      155      160
313 Ser Pro Ala Ala Thr Ala Ser Asp Pro Ala Gly Pro Pro Pro Leu Pro
314      165      170      175
317 Leu Pro Gly Pro Pro Pro Leu Ala Pro Thr Ala Thr Ala Gly Thr Leu
318      180      185      190
321 Ala Ala Ser Glu Gly Arg Trp Lys Ser Met Arg Lys Ser Pro Leu Gly
322      195      200      205
325 Gly Gly Gly Gly Ser Gly Ala Ser Ser Gln Ala Ala Cys Leu Lys Gln
326      210      215      220
329 Ile Leu Leu Leu Gln Leu Asp Leu Ile Glu Gln Gln Gln Gln Gln Leu
330 225      230      235      240
333 Gln Ala Lys Glu Lys Glu Ile Glu Glu Leu Lys Ser Glu Arg Asp Thr
334      245      250      255
337 Leu Leu Ala Arg Ile Glu Arg Met Glu Arg Arg Met Gln Leu Val Lys
338      260      265      270
341 Lys Asp Asn Glu Lys Glu Arg His Lys Leu Phe Gln Gly Tyr Glu Thr
342      275      280      285
345 Glu Glu Arg Glu Glu Thr Glu Leu Ser Glu Lys Ile Lys Leu Glu Cys
346      290      295      300
349 Gln Pro Glu Leu Ser Glu Thr Ser Gln Thr Leu Pro Pro Lys Pro Phe
350 305      310      315      320
353 Ser Cys Gly Arg Ser Gly Lys Gly His Lys Arg Lys Ser Pro Phe Gly
354      325      330      335
357 Ser Thr Glu Arg Lys Thr Pro Val Lys Lys Leu Ala Pro Glu Phe Ser
358      340      345      350
361 Lys Val Lys Thr Lys Thr Pro Lys His Ser Pro Ile Lys Glu Glu Pro
362      355      360      365
365 Cys Gly Ser Leu Ser Glu Thr Val Cys Lys Arg Glu Leu Arg Ser Gln
366      370      375      380
369 Glu Thr Pro Glu Lys Pro Arg Ser Ser Val Asp Thr Pro Pro Arg Leu
370 385      390      395      400
373 Ser Thr Pro Gln Lys Gly Pro Ser Thr His Pro Lys Glu Lys Ala Phe
374      405      410      415
377 Ser Ser Glu Ile Glu Asp Leu Pro Tyr Leu Ser Thr Thr Glu Met Tyr
378      420      425      430
381 Leu Cys Arg Trp His Gln Pro Pro Ser Pro Leu Pro Leu Arg Glu
382      435      440      445
385 Ser Ser Pro Lys Lys Glu Glu Thr Val Ala Arg Cys Leu Met Pro Ser
386      450      455      460
389 Ser Val Ala Gly Glu Thr Ser Val Leu Ala Val Pro Ser Trp Arg Asp

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/054,935

DATE: 02/14/2002

TIME: 16:05:10

Input Set : A:\16U 107 R1.ST25.txt

Output Set: N:\CRF3\02142002\J054935.raw

L:9 M:270 C: Current Application Number differs, Replaced Current Application No

L:9 M:271 C: Current Filing Date differs, Replaced Current Filing Date